

**À LA RECHERCHE DES RACINES PERDUES**  
**(IN SEARCH OF LOST ROOTS)**

MARCO ABATE

Dipartimento di Matematica

Università di Pisa

Largo Pontecorvo 5

56127 Pisa

Italia

Telefono: +39/050/2213.230

Fax: +39/050/2213.224

E-mail: [abate@dm.unipi.it](mailto:abate@dm.unipi.it)

*Talk given in the Congress*  
*“Matematica e Cultura 2012”*  
*on March 23, 2013*

A task traditionally associated to mathematics is the task of solving equations, and in particular polynomial equations: given a polynomial

$$p(x) = a_d x^d + \cdots + a_1 x + a_0,$$

where  $d$  is the *degree* of the polynomial and  $a_0, \dots, a_d$  are the *coefficients* of the polynomial, a *solution* or *root*<sup>1</sup> of the equation  $p(x) = 0$  is a number  $x_c$  such that the value of the polynomial  $p$  computed in  $x_c$  is equal to 0.

For polynomials (or, equivalently, equations) of low degree there are formulas for finding the roots starting from the coefficients. For instance, an equation  $ax^2 + bx + c = 0$  of degree two has (at most) two roots  $x_+$  and  $x_-$  given by the formula

and similar (though much more complicated) formulas are known for equations of degree three and four. On the other hand, the Norwegian mathematician Niels Henrik Abel in 1824 has shown that *no such formula can exist for equations of degree five or more*.

But wait a second. Any computer, given the right software (and enough time), is able to find solutions of equations of degree one thousand billions; so, if there are no formulas for solving the equation, how can it do it?

The solution (indeed...) of this conundrum lies in what one means by “having a formula”. A “formula” is a procedure (an algorithm) that starting from the coefficients of the equation produces roots of the equation. Abel’s result says that for equations of degree 5 or more there is no procedure *involving only arithmetic operations and extractions of square, cubic or, more generally,  $n$ -th roots* that works for all equations of the same degree; but it does not exclude the existence of other kinds of procedure.

To understand what alternative procedures might exist, let us give another look at the formula for solving equations of second degree. In that formula a square root appears<sup>2</sup>; but how does one compute the square root of a number? Well, there is an algorithm, known since the middle age (or at least since middle school), yielding arbitrarily good approximations of the square root, up to any prescribed decimal digit. And in general this is the most one can hope for: a square root usually is a irrational number, that is a decimal number with infinitely many digits not repeating in a periodic pattern, and thus it cannot be expressed exactly, with all its digits, in a finite amount of time. But for any practical<sup>3</sup> application knowing, say, the first one thousand digits of a square root is more than enough; and, by the same token, knowing the first one thousand digits of the solution of any equation should be enough.

---

<sup>1</sup> The use of the word “root” to denote a solution of an equation (see, e.g., [1, p.163]) goes back to the Arabic mathematician al-Khwārizmī, that in 830 AD wrote *Al-jabr*, the first modern algebra book. He thought of the variable  $x$  as the hidden source of the equation, as the root is the hidden source of a plant; and finding the solution as akin to bringing to light this source, that is to extracting the root from the soil.

<sup>2</sup> In a sense, this means that to solve the general quadratic equation is enough to be able to solve the very specific quadratic equation.

<sup>3</sup> And most unpractical applications too; for the remaining applications usually knowing the exact value of the square root is irrelevant, it suffices to know that it exists (and when it does).

Thus even in the quadratic case “solving” an equation means finding *approximations* of the roots up to a specified precision. This is exactly what (the software in) a computer does: it uses some algorithm for computing approximations to the roots of equations. There are several algorithms available for doing precisely this (see, e.g., [2] for a survey of the more important ones), some of them even faster than the classical formula in the quadratic case. The aim of this short note is to describe (part of) the story of one of the most well-known algorithms, that even though it has been around since the Seventeenth Century it is still being used and studied today, four centuries later: Newton’s method.

## 1. 1600: Newton (*et al.*)

This algorithm has been developed by Isaac Newton (yes, *the* Isaac Newton) in 1669 in a particular case (and expressed in a somewhat cumbersome way). Shortly later, John Wallis in 1685 and Joseph Raphson in 1690 gave a simpler description of the algorithm; and in 1740 Thomas Simpson devised the modern formulation we shall describe below, and noticed that Newton’s method can be also applied to solve some non-polynomial equations (see, e.g., [3] for more details).

The algorithm starts with a *seed*<sup>4</sup>, a first guess for the value of a root of the equation, chosen in any way you like; let us call this guess  $x_0$ . The idea is to apply to  $x_0$  a procedure producing a new guess closer to a root than  $x_0$ , and then repeat. Applying the procedure to  $x_0$  yields a new value  $x_1$  even closer to the root, applying it to  $x_1$  one gets  $x_2$  still closer, and so on; repeating the procedure enough times one gets an approximation of a root as good as needed. It is an example of *recursive* (or *iterative*) algorithm: the same steps are repeated over and over, using as input the output of the previous repetition, until the desired precision is reached.

The procedure for passing from  $x_n$  to  $x_{n+1}$  is easily described in graphical terms. Given the polynomial (or more general function)  $p$  and the seed  $x_0$ , the new approximation  $x_1$  is the  $x$ -coordinate of the intersection between the  $x$ -axis and the tangent line to the graph of  $p$  in the point  $(x_0, p(x_0))$ . In formula,

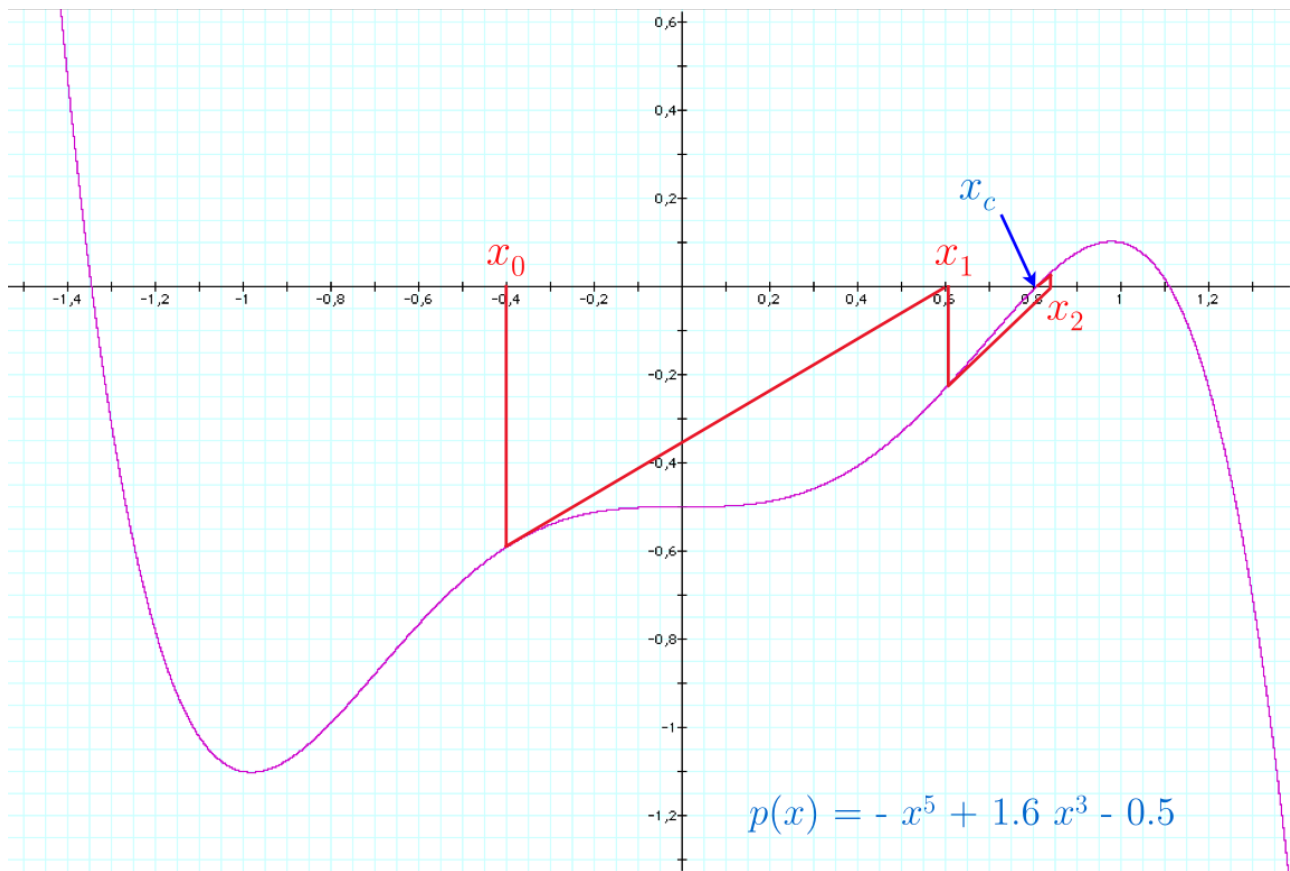
where  $p'(x)$  is the derivative of the polynomial  $p$  computed in the point  $x$ . Figure 1 shows how Newton’s method starting from the seed  $x_0$  (chosen at random) produces quite a good approximation of a root  $\alpha$  of the polynomial  $p$ . Indeed, two iterations are enough to get an approximation  $x_2$  correct up to the first decimal digit, three iterations are enough to get two decimal digits right (...) and four iterations yield four decimal digits (...).

Of course, to reliably use Newton’s method one must show that it works, that is that any sequence  $x_0, x_1, x_2, \dots$  of numbers obtained in this way actually gets arbitrarily close (in other words, *converges*) to a root of the original equation. And indeed Newton showed that if  $x_0$  is close enough to a root then the sequence produced by Newton’s method does converge to that root, so fast that in general the number of correct decimal digits doubles at each iteration (as happened in the previous example).

---

<sup>4</sup> Botanical metaphors abound in this subject...

So far so good: if the seed is close enough to a root, we have a fast and reliable procedure for finding that root. But wait: if we do not know the value of the root, *how can we possibly choose the seed to be close to a root whose location we do not know?*



**Figure 1:** An example of application of Newton's method.

## 2. 1800: Cayley

This is exactly the question posed in 1879 by Arthur Cayley (see [4]), an English mathematician. To try and find an answer to this question, he reformulated it as follows: given a polynomial  $p$ , let denote by  $f$  the rational function

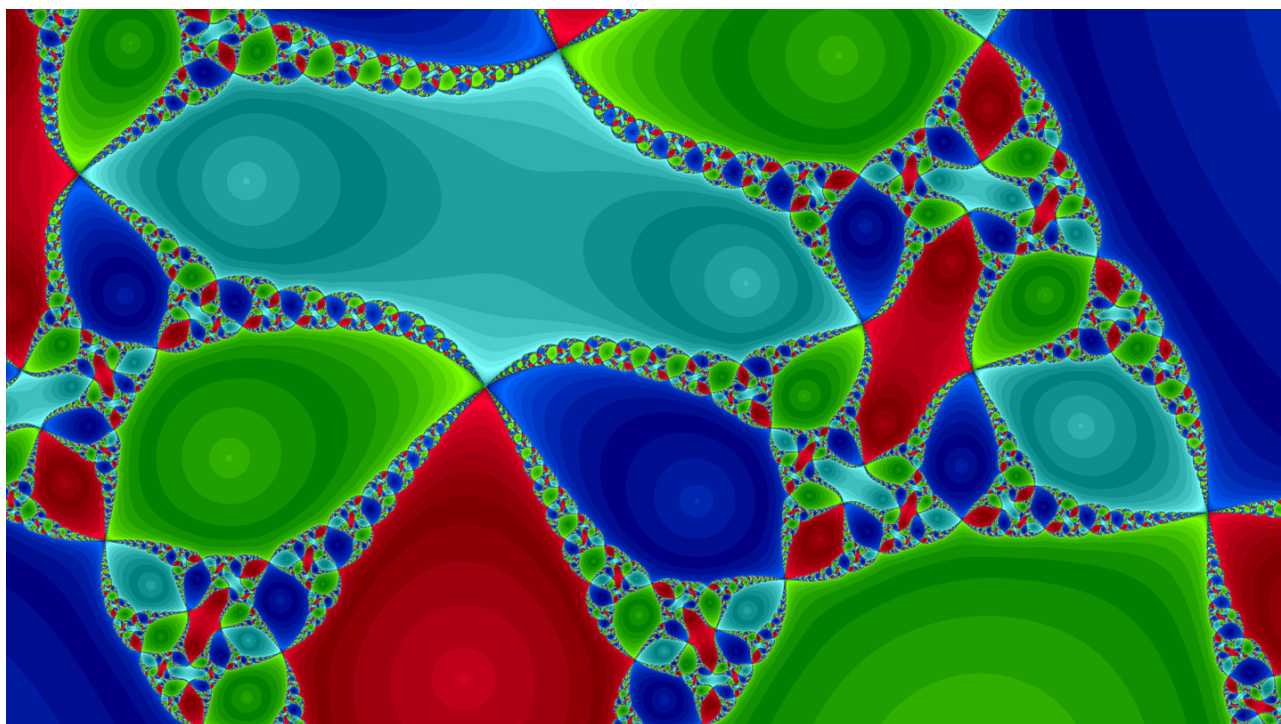
Given a number  $x_0$ , put  $x_1 = f(x_0)$  and, more generally, for every natural number  $n$ , put  $x_{n+1} = f(x_n)$ . In this way starting from the seed  $x_0$  we build a sequence  $(x_n)$ ; to be consistent with the botanical terminology used so far we say that  $x_0$  is *sprouting* the sequence. If this sequence converges to a number  $\alpha$  then necessarily one has  $f(\alpha) = \alpha$ , and this can happen if and only if  $\alpha$  is a root of the polynomial  $p$ , as explained by Newton. Therefore the question about when Newton's method works becomes: *for which values of the seed does the sequence converge?* If we call *basin of attraction* of a root  $\alpha$  the set of seeds sprouting sequences converging to  $\alpha$ , we would then like to know what is the union of the basins of attractions, and possibly we would like to know something about the geometry of these basins.

Since one has to start somewhere, Cayley began by considering the case of quadratic polynomials. He immediately noticed that the situation became much clearer working with *complex* numbers instead of real numbers, because a quadratic polynomial always has exactly two (possibly

coincident) complex roots, whereas it might not have real roots. Thus both the roots and the seed become points in the (complex) plane, and Cayley was able to prove that the bad seeds (that is, the seeds sprouting a sequence not converging to a root) are only the points of the *axis* of the segment connecting the two roots<sup>5</sup>; if the seed is chosen anywhere else Newton's method works. More precisely, the basin of attraction of a root is the open half-plane (bounded by the axis) containing the root; in other words, if the seed is on the left of the axis then Newton's method yields the left root, if it is on the right then Newton's method yields the right root. As a consequence, since the axis has zero area, choosing a seed at random on the plane we must be very unlucky to pick a point belonging to the axis, and thus applying Newton's method to a random seed will give a root.

So in the quadratic case Newton's method does not work for all seeds, but the set of bad seeds is so small (it has zero area) that if we pick a seed at random it will sprout a sequence converging to a root; in modern mathematical parlance, Newton's method in the quadratic case works *almost surely*. However, the quadratic case is a toy case: one can solve quadratic equations by using the classical formula, Newton's method is not needed. So Cayley tried to understand what happens for cubic polynomials, or for polynomials of higher degree—and he failed miserably. Axes, middle points or any other concept of Euclidean geometry seemed to be utterly useless in this setting. The situation was so complicated that Cayley was even unable to devise sensible questions (let alone answers) about the set of seeds for which Newton's method works...

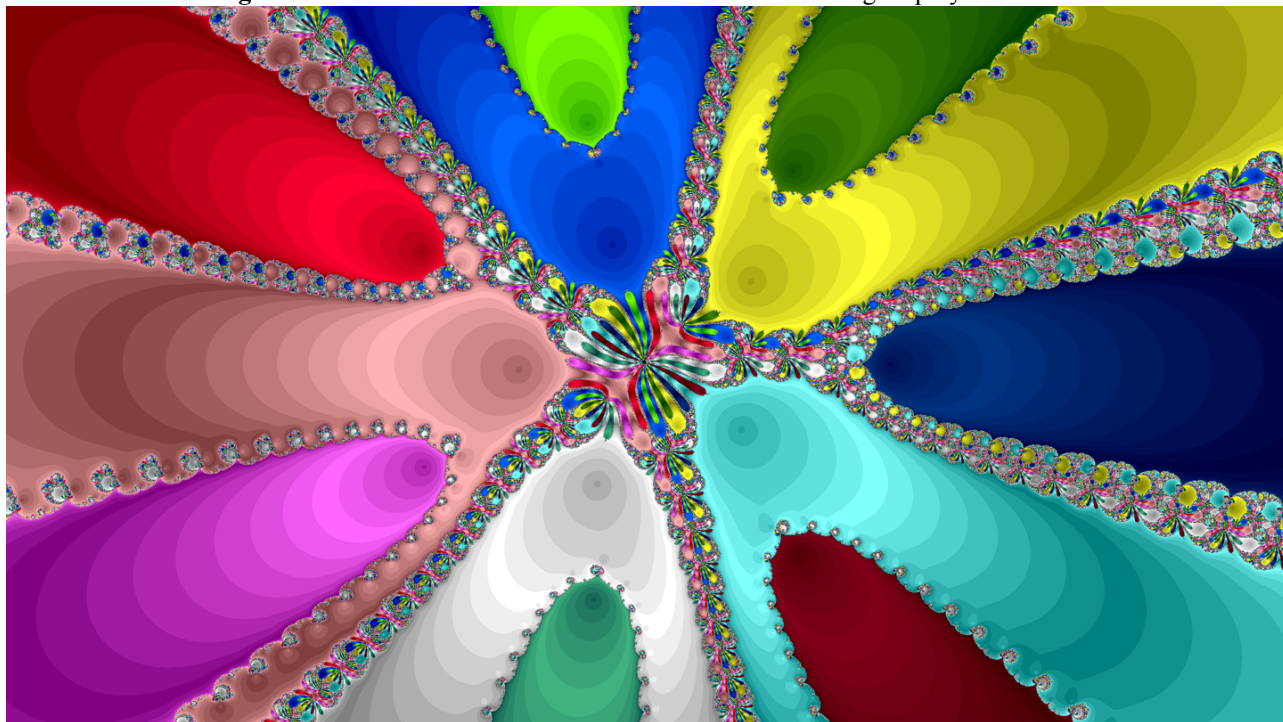
As it became clear a century later, Cayley cannot be blamed for this failure: the mathematical tools needed to deal with the problem were not available yet. And the geometry involved is highly non-trivial, as suggested by Figures 2 and 3, where different colors identify different basins of attraction; the analogous picture for a quadratic polynomial would have simply shown one half-plane in one color and the other one in a different color.




---

<sup>5</sup> If the roots are real, the only real bad seed is the middle point of the segment delimited by the two roots.

**Figure 2:** Basins of attraction of the four roots of a fourth degree polynomial.



**Figure 3:** Basins of attraction of the twelve roots of a twelfth degree polynomial.

### 3. 1900: Fatou (et al.)

During the latter half of the Nineteenth century and the beginning of the Twentieth century one branch of mathematics flourished (Complex Analysis of functions of one complex variable) and a new one was born (Dynamical Systems). And the deep new tools made available by these two branches allowed a French mathematician, Pierre Fatou, to start understanding the geometry of basins of attractions (see [5, 6, 7]).

To give an idea of the kind of results Fatou obtained, let us ask a seemingly unrelated question: is it possible to find in the plane two disjoint (open) sets having the same boundary? Well, that's easy enough: any closed curve without self-intersection subdivides the plane in two sets having the curve as common boundary<sup>6</sup>. For instance, a straight line (e.g., the axis of the segment connecting the two roots of a quadratic polynomial...) subdivides the plane in two open half-planes having the line as common boundary.

But what about three (or more) sets? Is it possible to find in the plane three or more disjoint sets all having the same boundary? Apparently not: as soon as we draw any kind of curve on the plane, no matter how short, we get two sides; how can that line be on the boundary of three or more disjoint sets? Clearly it cannot; and thus three disjoint sets with the same boundary cannot exist. Or can they?

The problem in this argument is that it is too naïve: the plane contains sets with boundaries so convoluted that they cannot be considered “curves”. Using a terminology that when Fatou was working was yet to be devised, there are sets with *fractal* boundaries. And basins of attraction are a primary source of examples of such sets.

Indeed, the surprising fact proved by Fatou is that *the  $n$  basins of attraction of the  $n$  roots of a polynomial of degree  $n$  all share the same common boundary*. The trick here is that (except in the

---

<sup>6</sup> This seemingly self-evident statement actually is a deep theorem of plane geometry known as *Jordan curve theorem*; see, e.g., [8] for a proof.

quadratic case...) a basin of attraction is not a single piece (is not connected, in modern mathematical terminology), but it is the union of infinitely many pieces all interconnected and touching all the infinitely many pieces of the other basins of attraction over and over in such a way that from any basin we can cross into any other basin at any point of the boundary.

For instance, around the blue set in the lower center of Figure 2 one can see a smattering of smaller sets of the other three colors (and some blue too), each one surrounded by even smaller sets of all four colors, and so on... the net result is that the blue basin (the union of all blue sets) touches the red, green and light blue basins in all points of its boundary, and the same can be said for the other basins. And Figure 3 shows the same phenomenon with 12 basins...

A very remarkable fact is that Fatou proved this statement in 1920, when no computer was around and he had no way of drawing, not even sketchily, these sets. He understood what was going on only using abstract mathematical tools, without relying on any kind of geometrical intuition (that furthermore pointed the other way, as we have seen); very remarkable indeed.

Fatou was able to say a lot more on the geometry of the basins of attraction of Newton's method (see, e.g., [9] for a modern survey of his results), but even he was not able to answer the question we are interested in: if we take a seed at random, does the sequence it sprouts converge to a root? In other words, does Newton's method almost surely work for polynomials of arbitrary degree?

#### 4. 2000: Hubbard (et al.)

Somewhat surprisingly, the answer is *no*, and it is so already for cubic polynomials. The first example showing what can go wrong was found in the 1970's, but possibly the easiest one is given by the polynomial  $p(z) = z^3 - 1$ . In this case Newton's method is expressed by the rational function

Figure 4 shows the three basins of attraction for Newton's methods, and a fourth black region. If the seed belongs to the black region, the sequence sprouted by the seed will eventually land in the central black region, and then it will keep jumping back and forth from that piece to the smaller black piece immediately to its right, without converging to any root (without converging anywhere, actually). For instance, if we get  $z_0 = 1$ ,  $z_1 = i$ ,  $z_2 = -1$ , and so on: the sequence keeps alternating 0 and 1, and neither 0 nor 1 are roots of the polynomial  $p$ . Since the black region has a positive area, choosing a seed at random there is a positive probability of picking it in the black region of bad seeds, where Newton's method does not work, and so there is the concrete danger of not being able to find some of the roots of the polynomial  $p$ .

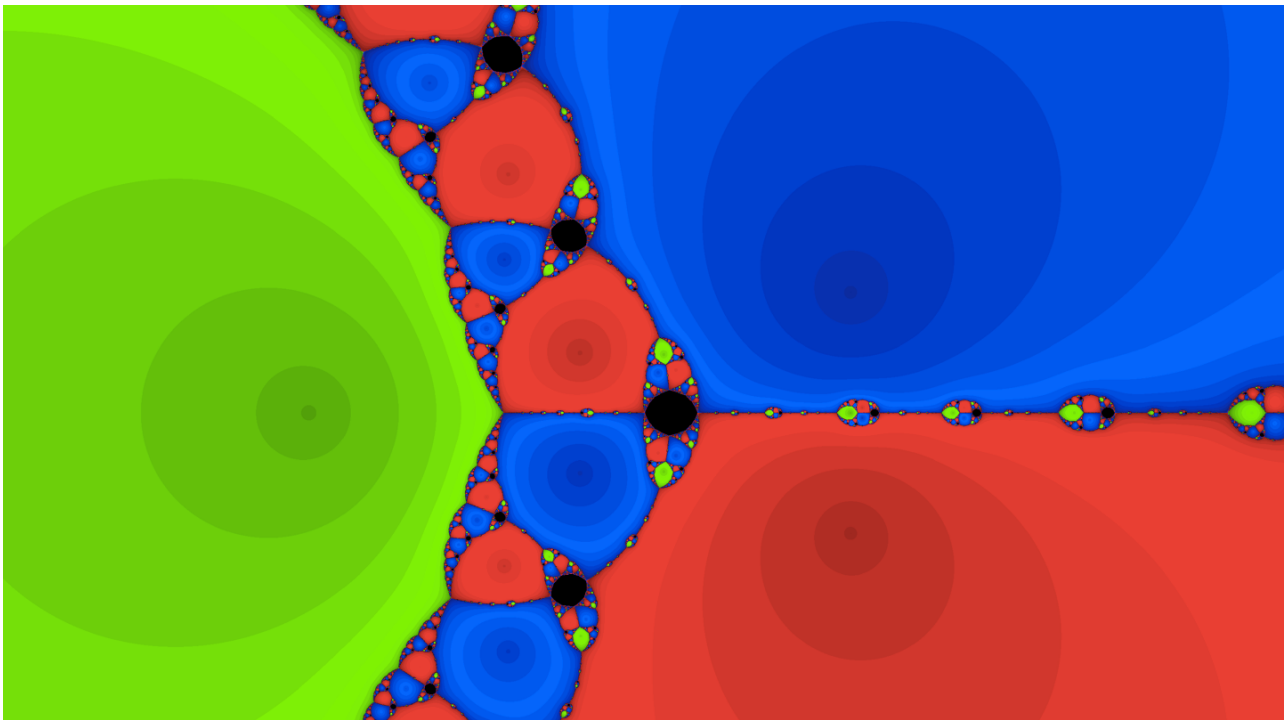
Is all lost, then? Should we forget Newton's method and recur to other methods for finding roots of polynomials? Possibly even more surprisingly, John Hubbard, Dierk Schleicher and Scott Sutherland in 2001 have shown (see [10]) that the answer also to this question is *no*: Newton's method *can* be used for finding the roots of polynomials of any degree; it suffices to choose the seeds in a smart way, and not at random.

More precisely, given a degree  $d$ , Hubbard, Schleicher and Sutherland have found an explicit way of building a set of points that can be used as seeds for finding *all* roots of *any* polynomial of degree  $d$ . In other words, if  $\alpha$  is a root of a polynomial  $p$  of degree  $d$ , we can always find in  $S$  at least one seed sprouting a sequence converging to  $\alpha$ .

The construction of the set  $S$  is a beautiful mixture of geometrical intuition and deep mathematics. Hubbard, Schleicher and Sutherland noticed (and proved!) that all basins of attraction for Newton's methods have "channels" going to infinity, that is strips of a definite width extending to infinity; furthermore, the complicated part of the basins is concentrated only in a bounded part of the plane and along the boundaries of the channels (see for instance Figure 3, where the channels are clearly



visible). Hubbard and collaborators then found a (by no means trivial) way for estimating both the diameter of the complicated part and the width of the channels; and they constructed  $S$  by picking points placed approximately along a circumference centered at the origin of radius large enough to be outside the complicated part, and distributed close enough to each other to hit all channels, thus ensuring the presence in  $S$  of at least one point of each basin.



**Figure 4:** Basins of attraction of the roots of a cubic polynomial, with a black region of bad seeds.

Another non trivial point here is how many points are in  $S$ ; if  $S$  contains too many point its use might be unpractical (nobody would try one million seeds for finding the three roots of a cubic polynomial). Now, to find all roots of a polynomial of degree  $d$  one clearly needs at least  $d$  seeds, one for each root. Finding a set of exactly  $d$  seeds able to generate all  $d$  roots of any polynomial of degree  $d$  might be impossible; but Hubbard and collaborators managed to use only slightly more than  $d$  points. Indeed, the set  $S$  they constructed contains approximately  $1.5d$  points. For instance,  $S$  contains 4 points (and not one million),  $S$  contains 9 points,  $S$  contains 15 points,  $S$  contains 59 punti,  $S$  contains 850 points, and so on.

From the point of view of computer programs, a growth of  $S$  is acceptable; the number of points to deal with is manageable even for polynomials of high degree. Furthermore, in 2012 Schleicher and others [11] have shown that Newton's method applied to  $S$  is at least as efficient as most of the other known algorithms for finding roots of polynomials.

Summing up, we now have an efficient way of using Newton's method for finding roots of polynomials. But this is just a small part of the story of Newton's method, and there is still a lot to understand and discover about it (see, e.g., [12] for a survey of recent results): after five centuries, there are still beautiful treasures to be found looking for lost roots.

## 5. *References*

- [1] W.W. Rouse Ball: **A short account of the history of mathematics**. Macmillan, London, 1893.



- [2] J.M. McNamee: **Numerical methods for roots of polynomials. Part I.** Elsevier, Amsterdam, 2007.
- [3] [http://en.wikipedia.org/wiki/Newton's\\_method](http://en.wikipedia.org/wiki/Newton's_method)
- [4] A. Cayley: *The Newton-Fourier Imaginary Problem.* Amer. J. Math. 2 (1879), 97.
- [5] P. Fatou: *Sur les equations fonctionnelles, I.* Bull. Soc. Math. France 47 (1919), 161–271.
- [6] P. Fatou: *Sur les equations fonctionnelles, II.* Bull. Soc. Math. France 48 (1920), 33–94.
- [7] P. Fatou: *Sur les equations fonctionnelles, III.* Bull. Soc. Math. France 48 (1920), 208–314.
- [8] M. Abate, F. Tovena: **Curves and surfaces.** Springer, Berlin, 2011.
- [9] J. Milnor: **Dynamics in one complex variable. Third edition.** Princeton University Press, Princeton NJ, 2006.
- [10] J.H. Hubbard, D. Schleicher, S. Sutherland: *How to find all roots of complex polynomials by Newton's method.* Invent. Math. 146 (2001), 1–33.
- [11] M. Aspenberg, T. Bilarev, D. Schleicher: *On the speed of convergence of Newton's method for complex polynomials.* Preprint, arXiv:1202.2475.
- [12] Johannes Rückert, *Rational and transcendental Newton maps.* In: **Holomorphic dynamics and renormalization**, Fields Inst. Commun. 53, Amer. Math. Soc., Providence, RI, 2008, pp. 197–211.